DOCUMENT RESUME

ED 435 114 EA 030 088

AUTHOR	Holdzkom, David
TITLE	Be Careful What You Ask for: The Impact of an Accountability System on Student Achievement, School Achievement and Teachers.
PUB DATE	1999-04-00
NOTE	14p.; Paper presented at the Annual Meeting of the American Educational Research Association (Montreal, Quebec, Canada, April 19-23, 1999).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Academic Achievement; *Academic Standards; *Accountability; Educational Assessment; Elementary Secondary Education; Evaluation Criteria; Needs Assessment; Problems; School Effectiveness
IDENTIFIERS	*North Carolina

ABSTRACT

This paper describes the development of teacher-accountability systems in North Carolina. It explains how state officials in the 1960s finally realized that the education of the citizenry was essential to the state's economic development. Initial efforts in school reform focused on the curriculum, but it soon become clear that simply writing new curriculum standards would have only minimal effects on what students actually learned. Attention then shifted to teachers, and it was decided that reform should be based on students' demonstrated knowledge. The report discusses the introduction of the end-of-grade (EOG) tests and the determination that students would be judged on a standard score that would determine the degree to which a school had met, failed to meet, or exceeded the expected growth for the students in that school. Schools were then assigned to four categories: meets expected growth, low-performing school, nonrecognition school, and meets exemplary growth. The article then explores the various problems this method engendered, such as concerns that the statistical complexity of the program was exacerbated by a lack of explanatory materials for the public and for teachers, and teachers' and schools' belief that truncating the curriculum was a reasonable strategy for meeting expectations. (Contains 15 references.) (RJM)

Reproductions supplied by EDRS are the best that can be made
from the original document.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
D Holdzkom
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

ED 435 114

BE CAREFUL WHAT YOU ASK FOR: THE IMPACT OF AN ACCOUNTABILITY SYSTEM ON STUDENT ACHIEVEMENT, SCHOOL ACHIEVEMENT AND TEACHERS

David Holdzkom

Durham Public Schools
PO Box 30002
Durham, North Carolina 27702

Paper presented at the Annual Meeting of the
American Educational Research Association
Montreal, Quebec, Canada
April 1999

EA 030 088

BE CAREFUL WHAT YOU ASK FOR: THE IMPACT OF AN ACCOUNTABILITY SYSTEM ON STUDENT ACHIEVEMENT, SCHOOL ACHIEVEMENT AND TEACHERS

David Holdzkom
Durham (NC) Public Schools

If innovation and experimentation were the hallmarks of education in the 1970's and 1980's, it can fairly be said that accountability is the battlecry of the late 1990s. While accountability is certainly not a new concept in education, it seems that, beginning in the early 1980's there was an underlying leitmotif of accountability, particularly as the innovators and experimenters appeared to be gaining the upper hand in education circles. Policy makers might read, and worry about, the jeremiads of the authors of *A Nation At Risk* and even enact policies intended to correct some of the shortcomings noted in that report, the accountability measures that were undertaken seemed doomed to fail.

In the mid-1980's there was a heavy emphasis on the development of teacher evaluation systems. This was not a new direction for education planners, and, in some ways represented an application of some rational management dicta. If there was a problem with the product (student learning), then the worker (the teacher) should be held accountable. However, the political realities of the time made it unlikely that any of these accountability systems, no matter how well conceived, would result in any long-term or permanent changes (Holdzkom and Brandt, 1995). There was little political stomach for holding the workers accountable.

The state of North Carolina could serve as an illustration of these issues and changes. Beginning as early as 1946, the State had attempted to create a teacher evaluation system that would insure that students learned more (Holdzkom, 1995; McCall, 1952). At least, the evaluation system was intended to "weed out the dead wood" among the teaching ranks. Thus, from the beginning, there was a natural audience whose interests would best be served by resisting the evaluation system: the teachers whose work was to be evaluated and whose livelihood was at stake. Since teachers--even without the benefit of a union, as is the case in North Carolina--are numerous and willing to influence electoral politics, it soon became clear that evaluating only the teachers' performance was unlikely to be a winning proposition for bringing about education improvements.

Certainly, education improvement could have been expected to be important for North Carolina in the 1950's and 1960s. Like other states of the Old South, North Carolina had never emphasized education much during its history. Reliant on an economy that was based on farming, fishing, and some factory work, only a weak connection could be seen between education and the

skills needed to earn a living (Hall et al., 1990). All that changed in the 1960s as policy makers in the state began to diversify the economy, seeking to encourage the development of service industry, particularly banking, insurance, and what came to be known as "high tech" industry. Cheap land, a promise of improved transportation, and tax incentives all proved attractive to officials of such industries as IBM, Glaxo, Borroughs Wellcome, and other leaders in research and technology, all of whom built major development laboratories in the state.

The problem of a lack of an educated workforce became obvious pretty quickly. The service industries and research corporations experienced difficulty with finding workers who had appropriate skills. Unfortunately, simply importing such workers also proved difficult since many employees were unwilling to relocate their families to an area where the schools suffered from a poor reputation. Ironically, the workers needed for the high tech industries provided precisely those who were most demanding of quality schools for their children. A system of worker rotations into and out of North Carolina did not solve the problem for corporate management. The only feasible solution appeared to be improving the schools. Political pressure began to be exerted on the state's General Assembly and school districts to bring about the changes that were perceived to lead to improved education.

Initially, efforts were made to reform curriculum, often allowing great latitude to school districts to adopt textbooks that were consonant with goals and objectives that were practically endorsed nation-wide. It soon became clear, however, that simply writing new curriculum standards would have only minimal effect on what students actually learned. Attention shifted to those people charged with delivering the curriculum. Again, identifying the "deadwood" and getting rid of it led to the development of teacher evaluation systems intended to ensure the quality of instruction being delivered in the state's schools. These efforts, and their predictable lack of success are described elsewhere (see, for example, Kuligowski, Holdzkom and French, 1993; Holdzkom, 1987; and Stacy, Holdzkom, and Kuligowski, 1989). While we need not take the time to describe the state's teacher evaluation system in detail, it should be noted that the evaluation system and a complementary career development program that offered salary incentives to higher performing teachers were supported by the school districts involved in early experimentation, but the system failed to take hold (Brandt et al., 1988). The state teachers' association complained that the evaluation system was overly mechanistic in its conception of the relationships among teachers, students, and learning, that the evaluation system did not hold the student responsible for his/her own learning, and that the emphasis on individual merit undermined the notion of team work that was essential for effective schools. Without reviewing the merits of any of these arguments, it is sufficient to note that the evaluation system continued to be employed (with varying degrees of

effectiveness) especially for the licensing of new teachers, while the incentive structure disappeared.

As the 1980s ended, the General Assembly, unwilling to abandon the notion of accountability, enacted programs that gave large degrees of latitude to teachers and principals to define the "improvements" in education for which they were willing to be accountable (NC General Assembly, 1989). Not surprisingly, many schools and districts proposed plans that would actually [illegible] about a minimum of improvement while avoiding charges that they were [illegible]operative in the effort to improve student learning (Holdzkom and [illegible]vski, 1993). This transition phase lasted for a few years before the [illegible]al Assembly grew frustrated by continuing to support programs that a[illegible]ared to have little impact on education improvement. Moreover, the rise of calls for "privatization" of education was exerting pressure to accept the fact that public schools couldn't change and that, therefore, investing in other arrangements--charter schools, vouchers, and similar plans--would at least offer parents more control over the quality of education delivered to students.

In an effort to de-fuse a political call for more schools of choice in the state, the State Board of Education, under the leadership of Dr. Jay Robinson, a former superintendent of Charlotte Mecklenburg Schools, launched a new accountability program predicated on the notion that schools could be classified as a function of their ability to provide a year's worth of academic growth for a year's worth of attendance (Tuttle, 1995). Launched as the *ABCs of Education*, his effort was possible because of the existence of a state-wide testing program that had been in place since school year 1992-93 (North Carolina Board of Education, 1995). Two separate, but related standards were set by which schools performance was to be measured: *a growth standard*, which calculated how much students had achieved during the year, and *the performance standard*, a description of the percent of students in the school at or above grade level.

The testing program, called the End of Grade (EOG) testing program, was initiated in concert with major curriculum revision in both reading and mathematics, which had been instituted in that school year. Thus, there was an almost perfect match between the goals and objectives of the curricula and the items included on the EOG tests. Originally intended to measure the degree of fidelity of implementation of the new curricula, the EOG tests were administered in three parallel forms. This was necessary because the number of academic objectives at any grade level made it impossible to test every student on the complete curriculum. However, by distributing objectives across the three forms, and then analyzing the achievement of the group, inferences about the quality of implementation of the curriculum at the grade and school level could reasonably be drawn. While differences across forms made individual performance analysis problematic, these difficulties disappeared at the level of the class.

Scores were reported for individual and groups (averages) using a developmental scale score running from 100 to 200, where 4 or 5 points represented the grade to grade differences in reading and 6-8 points represented grade to grade differences in mathematics (Pommerich et al., 1993)). At each grade level, score ranges were determined to fall into one of four Achievement Levels, with Levels III and IV considered to be "at or above" grade level. Indeed, Achievement Level III corresponded to about the 33rd state percentile, while Achievement Level IV corresponded to about the 75th state percentile, although these numbers varied slightly from grade to grade (NC DPI, 1998). It should be noted that these percentiles were associated with *individual* student's scores on a criterion-referenced test. Although normally associated with norm-referenced tests, the percentiles were calculated for this test, because, in the opinion of policy makers at the State Department of Public Instruction, parents would be able to understand percentiles, but would have difficulty with the concept of developmental scale scores.

Every student in grades 3 through 8 had been tested annually in both reading and mathematics, thus creating a huge database of student performance that could be analyzed in a number of different ways. (Obviously, there were some students who were exempted from testing, notably students of limited English proficiency or students in Exceptional Education programs whose IEPs specifically prohibited participation in the testing program.) Drawing on this data base of achievement, statisticians at the State Department of Public Instruction designed a program that predicted achievement for each grade at each school in the state. Essentially, this regressed present achievement on past achievement, thus eliminating criticisms that social conditions (poverty, parental education, etc.) had not been considered in predicting growth. This "growth" prediction was calculated by taking into account the average growth from grade to grade in the state as a whole and then modifying this gain by calculating a factor for "true proficiency" (the degree to which any group was actually above or below the state average growth) and a factor to take into account regression to the mean. By converting each grade's achievement gain score to a standard score, the standard scores for each grade could be combined to determine the degree to which a school had met, failed to meet, or exceeded the expected growth for the students in that school for any given year. Because each grade group's achievement in any given year could be different from the gain made by a previous group, these predictions would need to be re-calculated each year.

In addition to the statistics used in the program, individual students had to meet three criteria to be included in the growth calculation. Each student had to have two sets of scores for both reading and mathematics: one set (for the prior academic year) represented the starting point for the growth calculation, while the second set--from the current year--represented the ending point. Also, each student included in the growth calculation had to be enrolled in the

school for at least one semester, specifically, for the 90 days preceding the administration of the test. Finally, each student included had to be following the state Standard Course of Study. Only students who met these criteria would be included in the "accountability measure of growth, although all students tested would be included in the "performance" standard. The performance standard of the ABCs program was given much less prominence in the accountability program than was the growth standard. The performance standard reported the percentage of students in the school who had earned test scores in Achievement Levels III and IV; that is, the percent of students in the school performing at or above grade level. The 50 percent mark was set as the minimum desired standard.

Four possible conditions were associated with the ABCs program as illustrated below:

	Performance Standard	
	>50% on grade level	<50% on grade level
Meet		
Fail		

Depending upon which condition a school fell into, it was assigned to a category:

1. **Meets Expected Growth**. Schools in this condition met their expected growth (on average, for all grades combined). During the second year of the program (but not during the first year) teachers and teacher assistants in these schools were awarded a bonus payment of $1000 and $500 respectively;

2. **Low-Performing School**. These schools failed to attain their expected growth *and* had fewer than 50% of students performing at or above grade level.

3. **Non recognition School.** These schools failed to meet their expected growth target, *but* had 50% or more of students performing at or above grade level.

4. **Meets Exemplary Growth**. Schools in this condition exceeded their expected growth by 10% or more. Teachers in these schools received bonuses of $1500, while teacher assistants received $750.

Clearly, then, the emphasis was on growth, rather than on performance, which only factored into the definition of schools in the negative ratings categories. Thus, an objective of many schools was met in the distribution of rewards. That is, it had been argued that schools serving large numbers of poor children, a majority of whom were not at grade level would not be able to quality for the rewards for teachers and others, since it could not be expected that the performance standard would be met, no matter how much growth was attained. By eliminating the performance standard in calculating the rewards, this argument was addressed, and all schools were , presumably, motivated to do better in the future than they had done in the past. Similarly, by not relying on performance alone, schools serving large numbers of advantaged students would be required to continue to ensure academic growth for these students.

During the first year of the program's implementation, the results were about what one might expect (NCDPI Website). The distribution of schools to categories is shown in the graphic below:

	Performance Standard	
	>50% on grade level	<50% on grade level
Meet	909 (531 Exem)	17 (2 exem)
Fail	583	123

ABCs Program Results for 1996-97

The fact that the largest percentage of schools were in the No Recognition category (more than 50% of students at or above grade level but

failed to meet expected growth) is both interesting and instructive. These schools should have been able to meet the growth condition, since so many of their students were prepared to learn at grade level. Two possible explanations for the large number of this school might be that either the teachers did not understand what they needed to do in order to be successful in terms of the ABCs and/or that the teachers expectations for the students (and therefore the amount and quality of work that they set) was, in fact, below the ability of the students to learn.

By contrast, a combined 926 schools met or exceeded their growth goals. Actually, there were 531 schools in the Exemplary Growth category (32.5% of all participating schools) and 395 schools in the Expected Growth category (24.2% of all schools). While the combined total exceeded the number of schools in the No Recognition category, neither of these categories exceeded this latter category. Only 7.5% of schools fell into the low-performing category. Given the fact that this program was sparked in large measure by the perception that many schools were failing to deliver the necessary instruction to students, this finding must have come as a surprise, at least in some quarters. During this first year, as has been mentioned, teachers and teacher assistants in the Exemplary Growth schools received cash bonuses.

The outcomes for the 1997-98 school year are displayed in the graphic below. Among all the K-8 schools in the state, almost 66% were awarded the Exemplary Growth designation (NCDPI, 1998). An additional 18% met their growth goals. Less than 1% of the schools were designated Low Performing, while about 15% were designated No Recognition (called "adequate performance" in the second year of the program.

	Performance Standard	
	>50% on grade level	<50% on grade level
Meet	1419 (1123 exem)	22 (9 exem)
Fail	262	15

ABCs Program Results 1997-98 Results

Obviously, in the second year of the program, the results were even more positive than they had been the year before. Or at least they appeared to be. However, a number of unintended consequences had manifested themselves. Seven of these are especially interesting, because of the impact on various participants in the education enterprise. are described below.

1. The statistical complexity of the program was exacerbated by a lack of explanatory materials for the public and for teachers.

While the State Department of Public Instruction had spent a great deal of energy and effort on describing the program before its implementation, conducting both in-put and awareness sessions across the state, there was virtually no effort made to describe the statistics, and their rationale, which underlay the program. Thus, many teachers continue to believe that the growth formula is "unfair" because it does not take into account the poverty of children served in some schools. This problem was compounded after the release of the first year of data by officials of the State Department who were quoted as saying that there was "no correlation" between student poverty and ABCs outcomes". Clearly there was a correlation, as several independent analyses showed. Nevertheless the State Department of Public Instruction lost an important opportunity to explain how they had worked to ensure that the program was fair to all participants.

Education of the general public was even less a priority. While the law that enacted the ABCs program specifically required Low Performing schools to send a letter explaining their status to all parents whose students were assigned to these schools, there was no corresponding requirement for other schools to explain the basis of a positive designation. One result of this was that poorly informed teachers and principals tried to explain how the program worked to parents. The results were predictable.

During the second year of the program, this misunderstanding led to a fire-storm of criticism from those on the political right who could not believe that such a large percentage of North Carolina schools were "exemplary". Actually, of course, the designation spoke to "exemplary growth" an important distinction, since a school could, theoretically, be an exemplary growth school in which less than 50% of students were performing on-grade level work. By now, however, the damage was done and there was no good resolution of the policy disagreement.

2. For many teachers and schools, truncating the curriculum appeared to be a reasonable strategy for meeting the ABCs expectation.

While the A in ABCs spoke to accountability, the B designated basic skills. Only results in reading, mathematics, and writing were considered in

calculating the ABCs results. Many teachers (and parents) interpreted this to mean that only these basic skills "counted" in their schools. Stories of teachers restricting lessons in science and social studies were endemic in the second year of the program. Many teachers of music and art, to name just two subjects, complained bitterly that they were expected to emphasize the basic skills in their classes, at the expense of the objectives unique to these disciplines.

While discussion of the high school accountability model is beyond the scope of this paper, it can reasonably be argued that this problem will be even greater in high schools, where the accountability model rests on the measured achievement in just 10 courses.

3. In combination, these two problems tended to increase cynicism among teachers.

While there have not been any published studies that would lead to this conclusion, conversation with teachers and principals makes it clear that there is a growing cynicism about the responsibility and authority of teachers and principals to make judgments about differential success of students. All that seems to count, for these people, are the ABCs results. One additional consequence of this policy is that many low-performing schools are finding it difficult to staff their schools adequately. Indeed, why would a skilled teacher accept assignment to a low-performing school, especially when both the financial rewards available to others as well as public opprobrium would overwhelm the teacher's efforts?

At a time when the literature on teacher empowerment and quality management principals suggests that teachers should be made more powerful in decision-making, especially where the education of their students is concerned (see, for example, National Board for Professional Teaching Standards, 1989), the accountability program is undercutting this sense of efficacy by relying exclusively on test results and growth predictions that result from the application of a statistical procedure that teachers and others understand poorly if at all.

4. The accountability program masks at least some intractable problems.

One of the chief efforts to make the accountability program fair was the exclusion of students who had not been enrolled in a school for at least one semester. The students who are, thus, excluded are the very students whose life chances are most at risk. In urban environments, frequent change of residence (and consequently of school assignment) is most likely experienced by children whose families are least stable. These are the students who have

traditionally been least well served by their schools. Yet no accountability for their learning is included for any school.

In addition to this problem, there is another difficulty that results from the decision to combine outcomes for all grades into a single indicator. It is possible for relatively poor performance in one grade level to be masked by higher than expected performance in another grade level. Similarly, high performance in reading, for example, could mask poor performance in mathematics because all the indicators for a given school are combined into a single indicator number. Thus, the public may come to perceive as "an exemplary growth school" one in which students actually are not learning mathematics as well as they should be. Given the fact that elementary teachers are often well trained as reading teachers and relatively poorly in mathematics makes this a very real possibility that does not serve the public's right to know very meaningful.

5. *The use of team rewards can serve as a disincentive for school principals to ensure that poorly performing teachers are removed/improved or can lead teacher frustration.*

Again, this problem is related to the fact that all outcomes are combined into a single indicator. In at least one case, the "exemplary growth" designation rested on the work of one grade-level outcome. One grade had enough "excess" points to cover the deficiencies of points in all other grades. Thus, based on the work of this single grade level, all teachers and teacher assistants in the school received the bonus payments. When this was pointed out in a faculty meeting, the response essentially was that, since the accountability system was really unfair anyway, this apparent injustice didn't matter. In any event, the fact in many schools is that teachers do not work as a team in many important areas and certainly, teachers are not accustomed to policing their own ranks, no matter how much the literature on professionalism may advocate this.

Many teachers are willing to be held accountable for their own work, but are reluctant to be responsible for the work of their colleagues, over whom they have little influence in any case. While, when examined from this positive point of view, this may seem a small problem, what happens if we take the opposite case? That is, what happens if we look at a successful teacher whose work is hidden by the fact that many of his/her colleagues are unable to attain the level of success that is needed for the school to be designated "expected growth"? A strong teacher would be wise to abandon a school where the majority of teachers are unable to bring about enough learning to qualify for the rewards. While team rewards may be acceptable are team punishments?

REFERENCES

Brandt, R., Duke, D., French, R., and Iwanicki, E. (1988). *A Review with Recommendations of the North Carolina teacher Performance Appraisal Instrument*, Raleigh, NC: North Carolina General Assembly.

Hall, J., Leloudis, J., Korstad, R., Murphy, M., Jones, L., and Daly, C. (1990). *Like a Family: the Making of a Southern Cotton Mill World.* Chapel Hill, NC: University of North Carolina Press.

Holdzkom, D. (1995). Politics *vs.* Technique, with an Emphasis on *Versus.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Holdzkom, D. and Brandt, R. (1995). From Accountability to Professional Development. in Daniel Duke, ed., *Teacher Evaluation Policy*. Albany, NY: SUNY Press.

Kuligowski, B., Holdzkom, D., and French, R. (1993) Teacher Performance Evaluation in the Southern States: Forms and Functions. *Journal of Personnel Evaluation in Education 6*: 335-358.

McCall, W. (1952). *Measurement of Teacher Merit.* Raleigh, NC: North Carolina Department of Public Instruction.

National Board for Professional Teaching Standards. (1989). *Toward High and Rigorous Standards for the Teaching Profession: A Summary.* New York: Author.

North Carolina Department of Public Instruction. (1996). *Information Sessions on ABCs.* Raleigh, NC: Author.

North Carolina Department of Public Instruction (1997). *A Report Card for the ABCs of Public Education. Volume 1.* Raleigh, NC: Author.

North Carolina Department of Public Instruction (1998). *A Report Card for the ABCs of Public Education. Volume 1.* Raleigh, NC: Author.

North Carolina Department of Public Instruction. (1998). *State Norms for the North Carolina Testing Program.* Raleigh, NC: Author.

North Carolina Senate. (1989). *Senate Bill 2: The School Improvement and Accountability Act.* Raleigh, NC: Author.

Pommerich, M., Billeaud, K., Williams, V., and Thissen, D. (1993). *User's Guide for the North Carolina End of Grade Tests.* Raleigh, NC: North Carolina Department of Public Instruction.

Stacey, D., Holdzkom, D., and Kuligowski, B. (1989). Effectiveness of the North Carolina Teacher Performance Appraisal System. *Journal of Personnel Evaluation in Education 3:* 79-106.

Tuttle, S. (1995). Breaking Up the Public Schools. *North Carolina, Vol. 53, No. 6:* 36.